(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 972 836 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
19.01.2000 Bulletin 2000/03

(21) Application number: **99110008.2**

(22) Date of filing: **21.05.1999**

(51) Int. Cl.[7]: **C12N 15/55**, C12N 9/22, C12N 15/63, C12N 5/10, C12P 21/00

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **22.05.1998 JP 14186198**

(71) Applicant:
THE INSTITUTE OF PHYSICAL & CHEMICAL RESEARCH
Wako-shi, Saitama 351-0198 (JP)

(72) Inventors:
- **Morishima, Nobuhiro,**
  c/o The Institute of Physical
  Wako-shi, Saitama 351-0198 (JP)
- **Shibata, Takehiko,**
  c/o The Institute of Physical
  Wako-shi, Saitama 351-0198 (JP)
- **Mizumura, Hikaru**
  Fujimi-shi, Saitama 354-0004 (JP)

(74) Representative:
VOSSIUS & PARTNER
Siebertstrasse 4
81675 München (DE)

(54) Endonuclease

(57) The present invention relates to a site-specific endonuclease which recognizes a specific nucleotide sequence, to a gene coding for the endonuclease, to a recombinant vector containing the gene, to a transformant containing the vector, and to a process for producing the endonuclease.

EP 0 972 836 A2

Printed by Xerox (UK) Business Services
2.16.7/3.6

Description

FIELD OF THE INVENTION

[0001] The present invention relates to a site-specific endonuclease which recognizes a specific nucleotide sequence, to a gene coding for the endonuclease, to a recombinant vector containing the gene, to a transformant containing the vector, and to a process for producing the endonuclease.

BACKGROUND OF THE INVENTION

[0002] Endonuclease is a nuclease (nucleic acid degrading enzyme) which hydrolyzes the phosphodiester bond of a polynucleotide chain. Endonuclease recognizes and binds to a specific nucleotide sequence along DNA molecules, whereby molecules within the recognition sequence is cut. Endonuclease is a requisite enzyme for today's advanced gene engineering techniques for cloning and analyzing genes.

[0003] A site-specific endonuclease Endo.*Sce*I (hereinafter, also referred to as "*Sce*I") from an eucaryotic microorganism (e.g., yeast) is known to be a heterodimer having subunits of 75 kDa and 50 kDa. The subunits of *Sce*I as well as genes encoding the subunits have been cloned, and the nucleotide sequences thereof have been determined (for 75 kDa subunit, see Morishima, N. et al., J. Biol. Chem. 265, 15189-15197 (1990) and for 50 kDa subunit, see JP-B-7-77556).

[0004] In order to widely utilize the above-described endonuclease for artificially modifying a biochemical agent, a gene or the like, the endonuclease needs to be mass-produced with a gene expression system. The endonuclease does not function unless it recognizes a specific nucleotide sequence, i.e., the endonuclease needs to be specific to the nucleotide sequence to be recognized.

[0005] The 50 kDa subunit of the above-described endonuclease *Sce*I is encoded by mitochondrial genomes of yeast (*Saccharomyces cerevisiae*). A gene of a mitochondrial genome of yeast contains codons unique to mitochondria which are different from amino acid codons (universal codons) used in gene expression systems from organisms generally used for mass expression of protein (*E.coli*, baculovirus, yeast, etc.). If this gene of the mitochondrial genome is directly used, the protein expression system hardly produces a protein of an original amino acid sequence. For example, while TGA is a stop codon as a universal codon, it is a different codon coding for other amino acid (Trp) in mitochondria. A gene may be normally expressed in mitochondria but expression of the same gene may not result in a complete protein in a general expression system such as *E.coli* due to incomplete translation caused by the stop codon.

SUMMARY OF THE INVENTION

[0006] The present invention aims at providing a site-specific endonuclease which recognizes a specific nucleotide sequence, to a gene coding for the endonuclease, to a recombinant vector containing the gene, to a transformant containing the vector, and to a process for producing the endonuclease.

[0007] The present inventors have gone through intensive studies to solve the above-described problems. As a result, they succeeded in producing a modified endonuclease capable of recognizing and cleaving a specific nucleotide sequence by substituting, in a gene encoding an amino acid sequence of the smaller subunit of an endonuclease from yeast, codons unique to mitochondria with universal codons, and in mass-expressing the endonuclease, whereby the present invention was accomplished.

[0008] Accordingly, the present invention relates to an endonuclease capable of recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof that comprises the sequence information necessary to be specifically recognised and/or cleaved by the endonuclease of the invention. When in the following reference is made to a fragment of the 26 nucleotide sequence mentioned above, it is to be understood that said fragment confers the ability to be specifically recognised and/or cleaved by the endonuclease of the invention. The person skilled in the art knows how to determine said functions without further ado, for example, by applying methodology involving gel shift assays.

[0009] Further, the present invention relates to a recombinant protein of either (a) or (b):

(a) a protein comprising the amino acid sequence represented by SEQ ID NO:3; or
(b) a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the protein comprising at least one deletion, substitution or addition of amino acid in the amino acid sequence represented by SEQ ID NO:3.

[0010] Moreover, the present invention relates to a gene encoding the recombinant protein of either (a) or (b):

(a) a protein comprising the amino acid sequence represented by SEQ ID NO:3; or

(b) a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the protein comprising at least one deletion, substitution or addition of amino acid in the amino acid sequence represented by SEQ ID NO:3.

[0011] In addition, the present invention relates to a gene containing DNA of either (c) or (d):

(c) DNA comprising the nucleotide sequence represented by SEQ ID NO:2; or

(d) DNA encoding a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the DNA being capable of hybridizing with DNA which comprises the nucleotide sequence represented by SEQ ID NO:2 under stringent conditions.

[0012] Furthermore, the present invention relates to a recombinant vector comprising the above-described gene.

[0013] Additionally, the present invention relates to a transformant comprising the above-described recombinant vector.

[0014] Moreover, the present invention relates to a process for producing the endonuclease, comprising the steps of: culturing the above-described transformant; and recovering from the culture an endonuclease capable of recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof.

[0015] The present invention also relates to an endonuclease produced by the process of the present invention.

[0016] Finally, the present invention relates to a kit comprising the endonuclease and/or the recombinant protein and/or the gene and/or the recombinant vector and/or the transformant of the present invention.

[0017] The components of the kit of the invention may be packaged in containers such as vials, optionally in buffers and/or solutions. If appropriate, one or more of said components may be packaged in one and the same container.

[0018] This specification includes part or all of the contents as disclosed in the specification and/or drawings of Japanese Patent Application No. 10-141861 which is a priority document of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019]

Fig. 1 shows amino acid sequences of an endonuclease before and after the modification;

Fig. 2 shows the steps for constructing plasmid pY673L;

Fig. 3 shows the steps for constructing plasmids pEN1.7 and pEN0.5;

Fig. 4 shows the nucleotide sequence of 50 kDa subunit gene of *Sce*I which has been modified to conform the universal code;

Figs. 5A and 5B are photographs of electrophoresis showing sequence-specific endonuclease activities of the 50 kDa subunit of the modified *Sce*I;

Figs. 6A and 6B show substitution site of the 50 kDa subunit from *Saccharomyces uvarum* and oligonucleotides used for the substitution; and

Figs. 7A and 7B are photographs of electrophoresis showing sequence-specific endonuclease activities of the 50 kDa subunit from *Saccharomyces uvarum*.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0020] Hereinafter, the present invention will be described in more detail.

[0021] The present invention aims at mass-expressing mitochondrial genome DNA encoding the smaller subunit (50 kDa) of an endonuclease from yeast by using a protein expression system such as *E.coli* or yeast. In accomplishing this aim, the present invention modifies, in a gene coding for an amino acid sequence of the smaller subunit, codons unique to mitochondria into universal codons. The present invention relates to such a modified smaller subunit capable of recognizing and cleaving 26 base pairs of the specific nucleotide sequence.

[0022] An endonuclease of the present invention (i.e., the 50 kDa subunit of an endonuclease from yeast; hereinafter also referred to as "Endo.*Sce*I 50 kDa") is prepared as follows.

(1) Designing mutated amino acid and introducing mutation

[0023] According to the present invention, the smaller subunit of endonuclease *Sce*I from *Saccharomyces cerevisiae* or the smaller subunit of endonuclease *Suv*I from *Saccharomyces uvarum* is used as a target for introducing a mutation. The smaller subunits of both *Sce*I and *Suv*I have molecular weights of 50 kDa. However, they differ from each other for having 2 different amino acids (Fig. 6A).

[0024] The gene coding for the subunit (50 kDa) of *Sce*I (hereinafter, referred to as "ENS2") is encoded by a mitochondrial genome, and thus contains genetic codes unique to mitochondria (Table 1).

Table 1

Difference between mitochondrial code and universal code		
	Amino acid to be translated	
Codon	Universal code	Mitochondrial code
TGA	STOP	Trp
ATA	Ile	Met
CTA or CTT	Leu	Thr

[0025] The nucleotide sequence of ENS2 is known (JP-B-7-77556; Nakagawa, K., Morishima, N., and Shibata, T., J. Biol. Chem. 266, 1977-1984 (1991)). When ENS2 is expressed in a general expression system such as *E.coli* according to the universal code, the translation is interrupted at TGA where it is read as a stop codon as can be appreciated from Table 1 (for example, ENS2 described in JP-B-7-77556 includes a stop codon TGA at nucleotides 97-99). While ATA is read as Ile according to the universal code, it is read as Met according to the mitochondrial code.

[0026] In order to construct a normal mass-expression system for ENS2, it is necessary to modify the genetic code of ENS2 such that the amino acid sequence obtained upon expression in a general expression system (e.g., *E.coli*) is identical to an amino acid sequence as expressed in a mitochondrial expression system. Thus, according to the present invention, a codon for Trp (TGA) according to the mitochondrial code (Table 1) is substituted for a codon (TGG) that will be translated into Trp according to the universal code. Such substitution is also applied to ATA, and CTA and CTT that are translated into Ile and Leu, respectively, according to the universal code (Table 1). There is no need of substituting other degenerating codons which code for Ile or Leu according to the universal code.

[0027] Basically, there are 37 amino acids as the candidates for modification within the amino acid sequence of the smaller subunit of *Sce*I (476 amino acids). Their positions are shown in Fig. 1 and Table 2. As to the endonuclease from *Saccharomyces uvarum* (*Suv*I), Gly at 217 and Asn at 346 (Fig. 1) are additionally substituted for Lys and Asp, respectively, so that a total of 39 amino acids are modified. It is not necessary to substitute all of the above 37 or 39 amino acids. The number of substitution may be 36, or 35 or less. Translations into codes unique to mitochondria may not be complete as long as the 26 nucleotides (SEQ ID NO:1) mentioned later are recognized by the endonuclease. The positions of substitution are summarized in Table 2 below.

Table 2

Position of substitution	Amino acid to be translated before modification	Amino acid to be translated after modification
33, 54, 247, 320, 433	STOP	Trp
35, 40, 45, 48, 65, 80, 86, 92, 107, 109, 111, 123, 154, 163, 168, 171, 177, 248, 313, 335, 347, 399, 465	Ile	Met
49, 99, 130, 135, 222, 267, 276, 395, 426	Leu	Thr

[0028] Substitution of the amino acids is conducted by substituting the nucleotide sequence of the gene encoding the amino acids for another nucleotide sequence (site-directed mutagenesis). Examples of mutagenesis include but not limited to the site-directed mutagenesis method by T. Kunkel (Kunkel, T.A., Proc. Natl. Acad. Sci. U.S.A. 82, 488-492 (1985)) and the Gapped duplex method. There is also a modified version of Kunkel method in which a maximum of 16 oligonucleotides for modification are simultaneously used (instead of using 1 or 2 oligonucleotides as the usual Kunkel method) to efficiently substitute a plurality of sites. According to the present invention, mutation can be introduced by

using a mutation introduction kit (for example, Mutant-K (Takara Shuzo, Co., Ltd.) or Mutant-G (Takara Shuzo, Co., Ltd.)) that utilizes site-directed mutagenesis, or by using LA PCR *in vitro* Mutagenesis series kit (Takara Shuzo, Co., Ltd.).

[0029] The oligonucleotides are designed and synthesized using the nucleotide sequence of ENS2 as a template (1431 base pairs: Nakagawa, K. et al., J. Biol. Chem. 266, 1977-1984 (1991); JP-B-7-77556) such that at least one base that is to be introduced with the mutation is flanked by about 8 to 30 bases (each oligonucleotide having a total of about 18 to 60 bases). The oligonucleotides can be obtained through chemical synthesis using a usual synthesizer.

(2) Preparation of endonuclease gene which has been introduced with mutation

[0030] Each of the oligonucleotides obtained as described in (1) above is phosphorylated at 5' end, synthesized using ENS2 as a template and subjected to ligation reactions. These reactions can be performed using T4 Polynucleotide Kinase (Takara Shuzo, Co., Ltd.), T4 DNA polymerase (Takara Shuzo, Co., Ltd.), T4 DNA ligase (Takara Shuzo, Co., Ltd.) or the like.

[0031] The nucleotide sequence of the thus-obtained DNA is determined. The determination of the nucleotide sequence may be conducted according to a known method such as Maxam-Gilbert chemical modification method or a dideoxynucleotide chain termination method using M13 phage. Generally, the sequence is determined by using an automatic nucleotide sequencer (e.g., ALF (Pharmacia), 373A DNA sequencer (Perkin-Elmer), etc.).

[0032] SEQ ID NOS:2 and 3 exemplify the nucleotide sequence of the gene of the present invention and the amino acid sequence of the endonuclease of the present invention, respectively. The endonuclease of the invention acquires the essential function of endonuclease *Sce*I or *Suv*I, i.e., the function of recognizing the consensus sequence "CANRYNNANNCYYGTTW" and a sequence similar thereto, by linking to the larger subunit of natural endonuclease. The endonuclease of the invention exerts the function of the smaller subunit of the natural endonuclease and can specifically recognize the 26 bases represented by "GCCCAGACATATCCCTGAATGATACC" (SEQ ID NO:1) or a fragment thereof that comprises the sequence information necessary to be specifically recognized and/or cleaved by the endonuclease of the invention.

[0033] The endonuclease of the present invention may include at least one mutation such as deletion, substitution, addition or the like of the amino acid as long as it can specifically recognize the above 26 bases (SEQ ID NO:1).

[0034] For example, the amino acid sequence represented by SEQ ID NO:3 may include deletion of at least one, preferably 1 to 10, more preferably 1 to 5 amino acids; addition of at least 1, preferably 1 to 10, more preferably 1 to 5 amino acids; or substitution of at least 1, preferably 1 to 10, more preferably 1 to 5 amino acids for another amino acids. The endonuclease of the invention does not have to include mutations of all of the above-described 37 or 39 amino acids as long as it can recognize the above 26 bases (SEQ ID NO:1).

[0035] The phrase "can recognize" as used herein refers to the function of the endonuclease of the invention to bind to a site of the 26 bases within the gene and to cleave the gene such that the 26 base pairs are separated into two fragments with staggered ends.

[0036] DNA that can hybridize with the above gene (SEQ ID NO:2) under stringent conditions may also be included in the gene of the present invention. The stringent conditions are, for example, a sodium concentration of 15 to 900 mM and a temperature of 37 to 70°C, preferably 68°C.

(3) Preparation and transformation of recombinant vector

(i) Preparation of recombinant vector

[0037] A recombinant vector of the invention may be obtained by ligating (inserting) the gene of the invention to (into) a suitable vector. The vector for inserting the gene of the invention is not limited to a specific one as long as it is replicable in a host cell. Examples of such vector include but not limited to plasmid DNA and phage DNA.

[0038] The plasmid DNA is, for example, plasmid from *E.coli* (e.g., pRSET, pTZ19R, pBR322, pBR325, pUC118, pUC119, etc.), plasmid from bacillus (e.g., pUB110, pTP5, etc.), or plasmid from yeast (e.g., YEp13, YEp24, YCp50, etc.). The phage DNA is, for example, λ phage or the like. Similarly, an animal virus vector such as a retrovirus or vaccinia virus vector, or an insect virus vector such as a baculovirus vector may also be used.

[0039] In order to insert the gene of the invention into the vector, the purified DNA is cleaved with a suitable restriction enzyme. Then, the cleaved fragment is inserted into the restriction site or a multicloning site of the suitable vector DNA.

[0040] The gene of the present invention should be integrated into the vector such that the gene is able to function. If desired, the vector of the invention may include, other than the gene of the invention and the promoter, for example, a *cis*-element (e.g., an enhancer), a splicing signal, a poly(A) tail signal, a selective marker, and a ribosome binding sequence (SD sequence). Examples of the selective marker include dihydrofolate reductase gene, ampicillin-resistant gene and neomycin-resistant gene.

(ii) Preparation of transformant

[0041] A transformant of the invention may be obtained by introducing the recombinant vector of the invention into a host cell in such a manner that the gene of interest is capable to be expressed. The host cell is not limited to a specific one as long as it can express the DNA of the present invention. Bacteria such as genus *Escherichia* (e.g., *Escherichia coli*), genus Bacillus (e.g., *Bacillus subtilis*), genus Pseudomonas (e.g,. *Pseudomonas putida*), genus *Rhizobium* (e.g., *Rhizobium meliloti*), yeast such as *Schizosaccharomyces pombe*, animal cells (e.g., COS and CHO cells), and insect cells (e.g., Sf9 and Sf21) are exemplified.

[0042] When a bacterium such as *E.coli* is used as the host, it is preferable that the recombinant vector of the present invention is capable of autonomous replication and includes a promoter, a ribosome binding sequence, the gene of the invention and a transcription termination sequence. The recombinant vector may also include a gene for controlling the promoter.

[0043] As the *E.coli*, *E. coli* K12 and DH1 are exemplified and as bacillus, *Bacillus subtilis* MI 114 and 207-21 are exemplified.

[0044] As the promoter, any promoter may be used as long as it can be expressed in the host cell like *E.coli*. For example, a promoter derived from *E.coli* or a phage, e.g., trp promoter, lac promoter, p_L promoter or p_R promoter, may be used. Artificially designed and modified promoter like tac promoter may also be used.

[0045] The recombinant vector may be introduced into the bacterium according to any method for introducing DNA into a bacterium. For example, calcium ion method (Cohen, S.N. et al., Proc. Natl. Acad. Sci., USA, 69: 2110-2114 (1972)) and an electroporation method may be employed.

[0046] An yeast such as *Saccharomyces cerevisiae*, Saccharomyces uvarum, *Schizosaccharomyces pombe* or *Pichia pastoris* may also be used as the host. In this case, the promoter may be any promoter that can be expressed in the yeast. Examples of such promoter include but not limited to gal1 promoter, gal10 promoter, heat shock protein promoter, MF 1 promoter, PHO5 promoter, PGK promoter, GAP promoter, ADH promoter and AOX1 promoter.

[0047] The recombinant vector may be introduced into the yeast by any method for introducing DNA into an yeast. For example, electroporation method (Becker, D.M. et al., Methods Enzymol., 194, 182-187 (1990)), spheroplast method (Hinnen, A. et al., Proc. Natl. Acad. Sci., USA, 75, 1929-1933 (1978)), or lithium acetate method (Itoh, H., J. Bacteriol., 153, 163-168 (1983)) may be employed.

[0048] An animal cell such as simian cell COS-7, Vero, Chinese hamster ovary cell (CHO cell), mouse L cell, rat GH3 or human FL cell may also be used as the host. As a promoter, for example, SR promoter, SV40 promoter, LTR promoter or CMV promoter may be used. Other than these promoters, for example, an early gene promoter of human cytomegalovirus may also be used.

[0049] The recombinant vector may be introduced into the animal cell, for example, by an electroporation method, a calcium phosphate method or a lipofection method.

[0050] An insect cell such as Sf9 cell, Sf21 cell or the like may also be used as the host. The recombinant vector may be introduced into the insect cell, for example, by a calcium phosphate method, a lipofection method or an electroporation method.

(5) Production of endonuclease

[0051] The endonuclease of the present invention may be obtained by culturing the above-described transformant, and recovering the endonuclease from the culture thereof. The term "culture" as used herein refers to a culture supernatant, a cultured cell or microbial cell, or a cell or microbial cell debris.

[0052] The transformant of the invention is cultured according to a general method used for culturing the host.

[0053] A medium for culturing the transformant obtained from a microorganism host such as *E.coli* or yeast may be either a natural or a synthetic medium as long as it contains carbon sources, nitrogen sources, inorganic salts and the like assimilable by the microorganism, and as long as it can efficiently culture the transformant.

[0054] As carbon sources, carbohydrate such as glucose, fructose, sucrose, starch; organic acids such as acetic acid, propionic acid; and alcohols such as ethanol and propanol may be used.

[0055] As nitrogen sources, ammonia; ammonium salts of inorganic or organic acids such as ammonium chloride, ammonium sulfate, ammonium acetate, ammonium phosphate; other nitrogen-containing compounds; Peptone; meat extract; corn steep liquor and the like may be used.

[0056] As inorganic substances, potassium dihydrogen phosphate, dipotassium hydrogen phosphate, magnesium phosphate, magnesium sulfate, sodium chloride, iron(II) sulfate, manganese sulfate, copper sulfate, calcium carbonate and the like may be used.

[0057] The cultivation is generally performed under aerobic conditions such as shaking or aeration agitating conditions at 37°C for 12 to 18 hours. During the cultivation, pH is maintained at 6.5 to 7.5, preferably 7.0. pH is regulated with an inorganic or organic acid, an alkali solution or the like.

[0058] During the cultivation, an antibiotic such as ampicillin, tetracycline or the like may be added to the medium if necessary.

[0059] When culturing a microorganism transformed with an expression vector using an inducible promoter, an inducer may be added to the medium at need. For example, when a microorganism transformed with an expression vector using Lac promoter or trp promoter is cultured, isopropyl 1-thio-β-D-galactoside (IPTG) or indoleacetic acid (IAA) may be added to the medium, respectively.

[0060] A transformant obtained by using an animal cell host may be cultured in a generally used medium such as RPMI1640 medium or DMEM medium, or a medium obtained by supplementing the generally used medium with fetal bovine serum and the like.

[0061] The cultivation is generally conducted under 5% CO_2 at 37°C for 1 to 3 days. During the cultivation, an antibiotic such as kanamycin, penicillin or the like may be added to the medium.

[0062] After the cultivation, where a microbial cell or another cell intracelluraly produced endonuclease of the invention, the endonuclease is extracted by disrupting the microbial cell or the other cell. Where a microbial cell or another cell extracellularly produced endonuclease of the invention, the culture solution is directly used. Alternatively, the microbial cell or the other cell is removed through centrifugation or the like before isolating and purifying the endonuclease of the invention from the culture through a general biochemical method for protein isolation and purification such as ammonium sulfate precipitation, gel chromatography, ion exchange chromatography, affinity chromatography, or a combination thereof.

EXAMPLES

[0063] Hereinafter, the present invention will be described in detail by way of examples which do not limit the technical scope of the present invention.

Example 1: Preparation of single-stranded template DNA encoding subunit of *Sce*I and containing deoxyuracil

[0064] Endo.*Sce*I 50 kDa subunit gene ENS2 (1431 base pairs; Nakagawa, K., Morishima, N., and Shibata, T., *J. Biol. Chem.* 266, 1977-1984 (1991)) was modified simultaneously within two regions of the gene, i.e., within the upstream moiety of 1.0 kilobase pair and the downstream moiety of 0.4 kilobase pair. For this purpose, *Eco*RI/*Eco*RI fragment (1671 base pairs) containing the full-length 50 kDa subunit gene (Nakagawa, K., Morishima, N., and Shibata, T. *J., Biol. Chem.* 266, 1977-1984 (1991)), and PstI/*Eco*RI fragment (534 base pairs) containing the downstream moiety of the 50 kDa subunit gene were separately cloned into phagemids pUC118 (Takara Shuzo, Co., Ltd.), and were named plasmids pEN1.7 and pEN0.5, respectively (Fig. 3). These phagemids were introduced into *E.coli* strains CJ236 (Takara Shuzo, Co., Ltd.) for transformation. The transformant *E.coli* strains were shake cultured at 37°C for 12 hours or longer so as to prepare preculture solutions. Twenty µl of each pre-culture solution was added to 2 ml of 2 x YT culture medium containing ampicillin (100 µg/ml) (Sambrook, J., Fritsch, E. F., and Maniatis; T., *Molecular Cloning: a laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)) and cultured at 37°C for 1 hour. To each medium, helper phage M13KO7 (2.0 x 10^{12} plaque-forming unit (pfu); Takara Shuzo, Co., Ltd.) was added to constitute 0.4% in volume of the medium, and the resultant was cultured at 37°C for 1 hour. Thereafter, kanamycin (100 µg/ml) was added and the resultant was cultured at 37°C for 14 hours. Phage particles released from *E.coli* into the media during the cultivation were recovered. Specifically, 1.5 ml of each culture solution was centrifuged (14,000 rpm, 5 min.) in a micro-centrifuge. 1.2 ml of the supernatant was collected and centrifuged under the same conditions to completely remove the cell, thereby obtaining 1.0 ml of the supernatant. Subsequent procedure for preparing the DNA was conducted according to DNA purification of bacteriophage M13 phage summarized in the experimental text of J. Sambrook et al. (Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: a laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)).

Example 2: Synthesis of oligonucleotides for Introducing site-directed mutation

(i) Synthesis of single strand

[0065] 50 kDa subunit gene ENS2 is encoded by mitochondria genome, and thus includes genetic code unique to mitochondria (Table 1). In order to subject ENS2 to a general mass-expression system, these unique codons must be replaced so as to correspond to the universal code. According to Example 2, the bases were substituted by using a modified method of T. Kunkel (Kunkel, T. A., *Proc. Natl. Acad. Sci. U.S.A.* 82, 488-492 (1985)). Whereas the general Kunkel method uses only one or two oligonucleotides for modification, the present method simultaneously used a maximum of 16 oligonucleotides for efficient substitution at multiple sites.

[0066] 33 oligonucleotides were designed which each contained the base to be substituted flanked by approximately

10 to 15 bases (Table 3).

Table 3

1.	AAAAGACTGGATTATAGAA (A)	(SEQ ID NO: 6)
2.	TGAATATATGTATAAATTT (A)	(SEQ ID NO: 7)
3.	TATTAAATGGGATAATAAT (A)	(SEQ ID NO: 8)
4.	TATTAGATATGTATTATAATG (A)	(SEQ ID NO: 9)
5.	TACACCTATGTCTAATAAA (A)	(SEQ ID NO:10)
6.	AAAATATTATGGATTATAAA (A)	(SEQ ID NO:11)
7.	TTTTATATTTTAAATAAAATGAAAATGGAAATGGATAATTATAATAATAATA (A) (A) (A)	(SEQ ID NO:12)
8.	AAAATATTATGAATAATTTAA (A)	(SEQ ID NO:13)
9.	ACTATCTAATATTGAAACTAATTTATCTAATAATTT (CT)	(SEQ ID NO:14)
10.	TTATTTAATGGATAAATAT (A)	(SEQ ID NO:15)
11.	ATAAATATATGAAATATTTAG (A)	(SEQ ID NO:16)
12.	ATAATTATATGTTTAATAATA (A)	(SEQ ID NO:17)
13.	GGAGGTATTACAATTACTAATCATGCTAATGAT (CTA)	(SEQ ID NO:18)
14.	TTTTAGTAGAAAAATGGATGGATACTTTAAAAGATA (A) (A)	(SEQ ID NO:19)
15.	AGCTAAAGAAAAGATTTTTACTAATATTTATAATAATTA (CT)	(SEQ ID NO:20)
16.	AAATATTATGGATATTAAA (C)	(SEQ ID NO:21)
17.	TAATTATTGGTTATCTGG (A)	(SEQ ID NO:22)
18.	ATCATCTATGTATAATCCT (A)	(SEQ ID NO:23)
19.	TTAAAAATATGAGACCTAG (A)	(SEQ ID NO:24)
20.	GATGAATTAATGAAATTTATTTA (A)	(SEQ ID NO:25)
21.	ATTAAATTTAGATTTAATACTTTTATTAAATCATATAAT (CTA)	(SEQ ID NO:26)
22.	TATAATAAATATATTAATATGCATAATGCACGTAAACC (A)	(SEQ ID NO:27)
23.	TAAATTTTTAATAAATAATATGACTTGTTTTATTAAATGgGA (ACTA)	(SEQ ID NO:28)
24.	AAGATTAATGAATTCAAAA (A)	(SEQ ID NO:29)
25.	GATTATAAATTATTATATACTTATTTTTATATTTTAAAT (CT)	(SEQ ID NO:30)
26.	gAATAATTTAAATTATAAAACTTCTAATATTGAAacTA	(SEQ ID NO:31)

	(CTA)	
27.	TTCTCTATTAATATTAAAACTAATTTAGCTAAAGAAA	(SEQ ID NO:32)
	(CT)	
28.	AAATTATTTACCAGAACTACTGATGAATTAATgAAATT	(SEQ ID NO:33)
	(CT)	
29.	CATATAATTGGAATAATAGA	(SEQ ID NO:34)
	(A)	
30	AATTTTTAATGAATAATATg	(SEQ ID NO:35)
	(A)	
31.	TTTAGATATGTTAAATATg	(SEQ ID NO:36)
	(A)	
32.	ATATgTTAAATATGATTCCTAATAA	(SEQ ID NO:37)
	(A)	
33.	CTGgATTATGGAATATGAAT	(SEQ ID NO:38)
	(A)	

[0067] In Table 3, the base(s) in parentheses underneath each sequence represent the original base(s) that was (were) substituted for the underlined base(s). The bases shown in small letters represent those which have already replaced the original oligonucleotide.

[0068] The lengths of the oligonucleotides vary within the range of 18 to 52 bases and they include mutation of 1 to a maximum of 4 residues. These oligonucleotides were used for substituting 50 base pairs of the 1431 bp 50 kDa subunit gene to modify 37 codons. 5' end of each oligonucleotide was phosphorylated so as to allow the DNA ligase reaction described later. The composition of the reaction mixture for the phosphorylation is shown below:

100 mM Tris-HCl (pH 8.0)
10 mM magnesium chloride
7 mM dithiothreitol
1 mM ATP, 1 μM oligonucleotide
T4 polynucleotide kinase (15 units)
Total amount 30 μl

[0069] The reaction mixture was subjected to phosphorylation reaction at 37°C for 15 min., and then the enzyme was inactivated through a treatment at 70°C for 10 min.

(ii) Synthesis of complementary strand

[0070] The oligonucleotides obtained in (i) were treated as follows to obtain double-strands. Compositions of the annealing buffer and the elongation reaction buffer are shown below:

Annealing buffer
200 mM Tris-HCl (pH 8.0)
100 mM magnesium chloride
500 mM sodium chloride

(continued)

Annealing buffer
10 mM dithiothreitol

Elongation reaction buffer
50 mM Tris-HCl (pH 8.0)
5 mM dithiothreitol
60 mM ammonium acetate
0.5 mM each of dNTPs (A, C, T, G)
5 mM magnesium chloride
1 mM nicotinamido adenine dinucleotide

[0071] Distilled water was added to 1 μl of the annealing buffer and 0.2 pmol of the single-stranded template DNA, resulting in a total amount of 10 μl. One μl of the solution was dispensed to be mixed with 1 μl of the phosphorylated oligonucleotide solution. The resultant mixture was left to stand at 65°C for 15 min. and then at 37°C for 15 min., whereby the oligonucleotide annealed to the single-stranded DNA. To the solution, 25 μl of the elongation buffer, 60 units of *E.coli* DNA ligase and 1 unit of T4 DNA polymerase were added and left to stand at 25°C for 2 hours so as to synthesize a complementary strand. Three μl of 0.2 M ethylene diamine tetra acetic acid tetrasodium salt (pH 8.0) was added to terminate the enzyme reaction, after which the enzyme was inactivated through treatment at 65°C for 5 min. This reaction solution was directly used for the subsequent transformation.

Example 3: Transformation

[0072] *E.coli* BMH71-18 mutS (Takara Shuzo, Co., Ltd.) was used such that the nucleotide sequence of the wild-type DNA strand (the single-stranded DNA prepared with CJ236) in the double-stranded plasmid DNA obtained through the complementary strand synthesis was substituted for a mutant type. In this *E.coli* strain, deoxyuracil contained in the single-stranded DNA prepared with CJ236 was hydrolyzed by enzyme uracil-DNA glycosylase and then synthesized again using the DNA strand containing the substituted base as a template (Lindahl, T., *Ann. Rev. Biochem.* 51, 61-87 (1982).

[0073] The whole reaction mixture with the synthesized complementary strand was added to 100 μl solution containing competent cell of BMH71-18 mutS. *E.coli* competent cell was prepared according to the method of H. Inoue et al. (Inoue, H., Nojima, H., and Okayama, H., *Gene* 96, 23-28 (1990)).

[0074] To the solution, a medium was added and left at 37°C for 1 hour. Then, 30 μl of helper phage (*supra*) was added and left to stand at 37°C for another 30 min. for infection to take place. 40 μl of culture solution of BMH71-18 intracellularly containing both the helper phage and the plasmid was fractionated and added to 2 ml of 2 x YT medium containing ampicillin (100 μg/ml) and kanamycin (100 μg/ml). The resultant medium was shake cultured at 37°C for 16 to 20 hours to produce a phage.

[0075] The microbial cell was removed through centrifugation (14,000 rpm, 5 min.). The supernatant was collected which contained the phage particle incorporating the single-stranded DNA of the plasmid with the substituted base. With 20 μl of the supernatant, 80 μl of strain MV1184 (Takara Shuzo, Co., Ltd.) which has been cultured for 12 hours or longer was mixed and left to stand at 37°C for 10 min. so as to inject the single-stranded DNA of the phage into the cell. The strain MV1184 containing the plasmid resulting from replication of the integrated single-stranded DNA was inoculated to an LB agar medium containing 100 μg/ml ampicillin (Sambrook, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning: a laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York (1989)) for selection. The nucleotide sequences of the 50 kDa subunit genes were analyzed for some clones with an automatic sequencer ALF (Pharmacia) to confirm the incorporation of the predetermined substitutions. The fluorescent primer for the analysis of the nucleotide sequence was purchased from Pharmacia (Uppsala, Sweden). The DNA sequencing reaction was based on the Sanger method (Sanger, F. et al., *Proc. Natl. Acad. Sci.*, 74, 5463-5467 (1977)) according to the protocol of Pharmacia.

[0076] For the starting material, plasmid pEN1.7, forty nucleotide substitutions were performed by 9 cycles of the

mutagenic process, while 10 nucleotide substitutions were performed for pEN0.5 by 4 cycles of such process.

[0077] Once all of the substitutions were confirmed, the upstream and downstream moieties were linked at the *Pst*I cleavage site, thereby obtaining a gene of the invention encoding the 50 kDa subunit with complete substitutions (Fig. 4, SEQ ID NO:2).

Example 4: Construction of expression plasmid

for 50 kDa subunit

[0078] For facilitating the linking between the modified 50 kDa subunit gene and the vector for inducing expression thereof, restriction sites were introduced into the 5' and 3' terminuses of the modified gene through polymerase chain reaction (PCR). The reaction was performed using Taq DNA polymerase (Takara Shuzo, Co., Ltd.) according to the protocol of the manufacturer. Sequences of the used primers are shown below.

5'-CCGGATCCATGAAAAAAC-3' (SEQ ID NO:4)
5'-GGGTCGACTTATTTAATGTATCC-3' (SEQ ID NO:5)

[0079] The underlined parts are the newly introduced *Bam*HI and *Sal*I recognition sequences. The reaction was performed through 25 cycles of: 94°C for 1 min.; 45°C for 2 min.; and 72°C for 3 min.

[0080] The DNA fragment (1447 base pairs) amplified by PCR was separated through agarose (0.8%) gel electrophoresis (Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: a laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)) and stained with ethidium bromide for confirmation. The fragment was recovered from the agarose gel using Geneclean kit (BIO101, California, USA).

[0081] The recovered DNA fragment was treated with *Bam*HI and *Sal*I, and subcloned into pRSET (Invitrogen Corp.) and pTZ19R (Pharmacia) to obtain pSC50 and pTZSC50, respectively. Plasmid pSC50 was used for inducing the expression. Plasmid pTZSC50 was subjected to DNA sequencing using fluorescent primer (*supra*) so as to confirm that no extra mutation had been introduced during the PCR.

Example 5: Induction of expression of 50 kDa subunit

[0082] Expression plasmid pSC50 was introduced into a competent cell of *E.coli* BL21 (DE3) pLysS (Invitrogen Corp.). The transformant cell was pre-cultured through shake cultivation in an LB liquid medium containing ampicillin (150 μg/ml) and chloramphenicol (34 μg/ml) (Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: a laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)) at 37°C overnight. The pre-cultured solution (4% in volume of a resultant culture) was centrifuged (2,500 x g, 10 min.) to recover the microbial cell. This precipitate was suspended in a small amount of fresh medium, which was then added to a liquid medium. Shake cultivation (37°C) was performed until the suspension level at 600 nanometers (nm) (OD600) of about 0.5 was obtained. Thereafter, shake cultivation was continued at 18°C. When the OD600 became about 0.8, isopropyl 1-thio-β-D-galactoside (IPTG) was added to the final concentration of 0.4 mM to initiate induction of expression of the 50 kDa subunit. After performing shake cultivation at 18°C for another 12 hours, *E.coli* was recovered through centrifugation, rapidly frozen with liquid nitrogen and stored at -80°C.

Example 6: Purification of 50 kDa subunit from *E.coli*

[0083] The microbial cell stored at -80°C was melted at room temperature. The subsequent treatments were conducted at 4°C or on ice. The microbial cell was suspended in Buffer A (20 mM Tris-HCl buffer (pH 8.0), 500 mM sodium chloride, 5 mM imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF; Sigma Aldrich Japan K.K., Tokyo, Japan), 0.1 % NP-40 (Nacalai Tesque, Inc., Kyoto, Japan)). The resultant suspension was rapidly frozen with liquid nitrogen and melted under running water to disrupt the *E.coli*. The suspension was treated for 5 times with an ultrasonicator (UR-200P, Tomy Seiko Co., Ltd., Tokyo, Japan) at a maximum output for 30 sec.

[0084] The treated solution was centrifuged (39,000 x g, 20 min.) at 4 °C. The obtained supernatant was filtrated through 0.45 μm Mylex filter (Millipore, Massachusetts, USA). The sample was placed in a column (Ø 10 mm, 2.0 ml) loaded with Probond Nickel Chelate Resin (Invitrogen Corp.) which had been equibralized with Buffer A. Then, the sample was washed with 20 ml of Buffer A (ten times the volume of the resin). After another washing with 12 ml Buffer A containing 60 mM imidazole (six times the volume of the resin), the resultant was subjected to gradient elution with 60-50 mM imidazole-containing Buffer A (total amount of 80 ml). The eluted fraction was subjected to sodium dodecyl sulfatepolyacrylamide gel electrophoresis (Laemmli, U. K. *Nature*, 227, 680-685 (1970)) and stained with Coomasie brilliant blue to confirm the presence of the 50 kDa subunit. The fraction containing the 50 kDa subunit was dialyzed

against Buffer B (20 mM Tris-HCl buffer (pH 7.5), 300 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid tetrasodium salt, 1 mM dithiothreitol). The purified protein was quantified using Protein assay agent (Bio-Rad, California, USA) according to the micro-assay method of the manufacturer. Bovine serum albumin solution (Sigma Aldrich Japan K.K.) was used as a standard protein. As a result, 300 μg of purified protein was obtained from 25 g (wet weight) of the microbial cell.

Example 7: Measurement of endonuclease activity

[0085] A substrate for measuring an endonuclease activity of the 50 kDa subunit was prepared as follows. An *Eco*RI/*Eco*RI fragment containing *oli*2 region on mitochondria DNA within which Endo.*Sce*I is known to cleave (1671 base pairs; Nakagawa, K. et al., *EMBO J.* 11, 2707-2715 (1992)) was subcloned into phagemid pUC119 (Takara Shuzo, Co., Ltd.), the resultant called pY673L (Fig. 2). Plasmid pBR322 (Takara Shuzo, Co., Ltd.) was used as a control DNA substrate. Plasmids pY673L and pBR322 were used to transform *E.coli*. Then, the plasmids were extracted from *E.coli* and highly purified using Qiagen column (Qiagen Japan, Tokyo, Japan).

[0086] The composition of the reaction solution used for measuring the endonuclease activity of the 50 kDa subunit is shown below:

50 mM Tris-HCl buffer (pH 8.0)
50 mM sodium chloride
10 mM magnesium chloride
1 mM dithiothreitol
25 ng substrate DNA (pY673L or pBR322 which has been linearized with restriction enzyme *Sca*I (Fig. 2))
0.4 to 60 ng 50 kDa subunit
Total volume 30 μl

[0087] After performing the DNA cleavage reaction at 37°C for 30 min., ethylenediaminetetraacetic acid tetrasodium salt and sodium dodecyl sulfate were added to final concentrations of 10 mM and 0.3%, respectively, to terminate the reaction. The cleaved DNA was subjected to 0.8% agarose electrophoresis either directly or after concentrating the DNA through phenol extraction and ethanol precipitation.

[0088] After the electrophoresis, the gel was stained with ethidium bromide (Sigma Aldrich Japan K.K.) or SYBR Green (Takara Shuzo, Co., Ltd.) to confirm cleavage of DNA. The DNA was detected using FMBIO Imaging device (Takara Shuzo, Co., Ltd.) to determine the DNA cleavage.

Example 8: Detection of sequence-specific endonuclease

[0089] The dimeric Endo.*Sce*I recognizes and cleaves *in vivo* and *in vitro* 26 base pairs similar to the consensus sequence within the *oli*2 gene region on the mitochondria DNA (Nakagawa, K., Morishima, N., and Shibata, T., *EMBO J.* 11, 2707-2715 (1992)) (Fig. 2). The purified 50 kDa subunit of Endo.*Sce*I cleaved a specific sequence (SEQ ID NO:1) by itself.

[0090] Specific cleavage of *oli*2 with the 50 kDa subunit using plasmid pY673L containing *oli*2 as the substrate was confirmed (Fig. 5A). Referring to Fig. 5A, Lanes 1 to 8 are the results obtained with 0.5, 1.0, 2.0, 4.0, 8.0, 16.0, 32.0 and 64.0 ng of the 50 kDa subunits, respectively. With 64 ng of 50 kDa subunit, 60% of pY673L (25 ng) in the reaction solution was sequence-specifically cleaved at 37°C within 30 min., whereby DNA fragments of 3.4 and 1.4 kilobases were detected.

[0091] DNA was not cleaved with the 50 kDa subunit using plasmid pBR322 as the substrate and no cleavage fragment was detected (Fig. 5B). Referring to Fig. 5B, Lanes 1 to 7 are the results obtained with 2.3, 4.5, 9.0, 18.0, 36.0, 72.0 and 144 ng of the 50 kDa subunits, respectively. When an excessive amount (200 ng) of the 50 kDa subunit was used, no cleavage was found with pBR322 or other DNAs (mitochondria DNA (80 kilobase pairs) from bud yeast strain, *E.coli* phage λ DNA (47 kilobase pairs), and bacillus phage φ105 DNA (38 kilobase pairs) which did not contain specific sequence (26 base pairs) within *oli*2 gene region).

Example 9: Mass production of 50 kDa subunit from *Saccharomyces uvarum* and detection of the activity thereof

[0092] Endo.*Suv*I 50 kDa subunit, a homologous protein of Endo.*Sce*I 50 kDa subunit from *Saccharomyces cerevisiae* is present in *Saccharomyces uvarum* (Nakagawa, K., Morishima, N., and Shibata, T., *J. Bio. Chem.* 266, 1977-1984 (1991)). Both subunits have 476 amino acid residues but there are two differences in amino acid level between them. The amino acid differences between Endo.*Sce*I and Endo.*Suv*I 50 kDa subunits are shown in Fig. 6A.

[0093] A mass-expression gene for Endo.*Suv*I 50 kDa subunit was prepared by introducing two additional modifications into the modified gene for Endo.*Sce*I 50 kDa subunit. The oligonucleotides used for the substitutions of the amino acids for Endo.*Suv*I 50 kDa subunit are shown in Fig. 6B. With reference to Fig. 6B, the bases in parentheses correspond to the nucleotide sequence of Endo.*Sce*I 50 kDa subunit. For this purpose, two oligonucleotides were newly synthesized to introduce mutations according to the gene modification method described above. The mutation was confirmed through DNA sequencing. The modified gene was subcloned into pRSET vector (Invitrogen Corp.), which was then introduced into *E.coli* BL21 (DE3) pLys by transformation method.

[0094] Endo.*Suv*I 50 kDa subunit was expressed and purified according to the method applied to Endo.*Sce*I 50 kDa subunit described above. The purified Endo.*Suv*I 50 kDa subunit was used to specifically cleave plasmid pY673L.

[0095] As a result, Endo.*Suv*I 50 kDa subunit was equally effective in sequence-specifically cleaving the Endo.*Sce*I 50 kDa subunit cleavage site on plasmid pY673L (Fig. 7A).

[0096] Meanwhile, Endo.*Suv*I 50 kDa subunit did not cleave plasmid pBR322 at all (Fig. 7B). Also, other DNAs (bud yeast mitochondria DNA, bacillus phage ϕ 105 λ DNA and *E.coli* phage λ DNA which did not contain *oli*2 gene region) were not cleaved by Endo.*Suv*I 50 kDa subunit.

[0097] In Figs. 7A and 7B, Lanes 1 to 7 are the results obtained with 2.3, 4.5, 9.0, 18.0, 36.0, 72.0 and 144 ng of 50 kDa subunits, respectively.

[0098] According to the present invention, a site-specific endonuclease capable of recognizing a specific nucleotide sequence, a gene encoding the endonuclease, a recombinant vector containing the gene, a transformant containing the vector, and a process for producing the endonuclease are provided. Since the endonuclease of the present invention is capable of recognizing a specific sequence of 26 bases, it is useful in the field of genetic engineering and biochemistry in modifying and mapping DNA for a wide application, i.e., plasmid to genome.

[0099] All publications, patents, and patent applications cited herein are incorporated herein by reference in their entirety.

[0100] The following are information on sequences described herein:

SEQUENCE LISTING

<110> The Institute of Physical and Chemical Research

<120> Endonuclease

<130> PH-651

<140>

<141>

<150> JP98/141861

<151> 1998-05-22

<160> 38

<170> PatentIn Ver. 2.0

<210> 1

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 1

gcccagacat atccctgaat gatacc 26

<210> 2

<211> 1431

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)..(1428)

<400> 2

atg aaa aaa caa aat tta aat tct att tta tta atg tat att aat tat 48
Met Lys Lys Gln Asn Leu Asn Ser Ile Leu Leu Met Tyr Ile Asn Tyr
1 5 10 15

att att aat tat ttt aat aat att cat aaa aat caa tta aaa aaa gac 96
Ile Ile Asn Tyr Phe Asn Asn Ile His Lys Asn Gln Leu Lys Lys Asp
20 25 30

tgg att atg gaa tat gaa tat atg tat aaa ttt tta atg aat aat atg 144
Trp Ile Met Glu Tyr Glu Tyr Met Tyr Lys Phe Leu Met Asn Asn Met
35 40 45

act tgt ttt att aaa tgg gat aat aat aaa att tta tta tta tta gat 192
Thr Cys Phe Ile Lys Trp Asp Asn Asn Lys Ile Leu Leu Leu Leu Asp
50 55 60

atg tat tat aat gta tta tat aac tat cat aaa caa cgt aca cct atg 240
Met Tyr Tyr Asn Val Leu Tyr Asn Tyr His Lys Gln Arg Thr Pro Met
65 70 75 80

tct aat aaa aga tta atg aat tca aaa aat att atg gat tat aaa tta 288
Ser Asn Lys Arg Leu Met Asn Ser Lys Asn Ile Met Asp Tyr Lys Leu
85 90 95

tta tat act tat ttt tat att tta aat aaa atg aaa atg gaa atg gat 336
Leu Tyr Thr Tyr Phe Tyr Ile Leu Asn Lys Met Lys Met Glu Met Asp
100 105 110

aat tat aat aat aat aat aat aat att tca tta aaa tat aat gaa tta 384
Asn Tyr Asn Asn Asn Asn Asn Asn Ile Ser Leu Lys Tyr Asn Glu Leu
115 120 125

tta aaa aat att atg aat aat tta aat tat aaa act tct aat att gaa 432

```
Leu Lys Asn Ile Met Asn Asn Leu Asn Tyr Lys Thr Ser Asn Ile Glu
    130                 135                 140
act aat tta tct aat aat ttt tat tta atg gat aaa tat tta att aat   480
Thr Asn Leu Ser Asn Asn Phe Tyr Leu Met Asp Lys Tyr Leu Ile Asn
145                 150                 155                 160
aaa tat atg aaa tat tta gat atg tta aat atg att cct aat aat tat   528
Lys Tyr Met Lys Tyr Leu Asp Met Leu Asn Met Ile Pro Asn Asn Tyr
                165                 170                 175
atg ttt aat aat att aat tat aaa ggt aaa tta aat att aaa aca gta   576
Met Phe Asn Asn Ile Asn Tyr Lys Gly Lys Leu Asn Ile Lys Thr Val
            180                 185                 190
tta gat tta aat aat aat gaa ttt tat gat tat tta tca ggg tta att   624
Leu Asp Leu Asn Asn Asn Glu Phe Tyr Asp Tyr Leu Ser Gly Leu Ile
        195                 200                 205
gaa ggt gat ggt tat att ggt cct gga ggt att aca att act aat cat   672
Glu Gly Asp Gly Tyr Ile Gly Pro Gly Gly Ile Thr Ile Thr Asn His
    210                 215                 220
gct aat gat gta tta aat act atc ttt att aat aaa aga att aaa aat   720
Ala Asn Asp Val Leu Asn Thr Ile Phe Ile Asn Lys Arg Ile Lys Asn
225                 230                 235                 240
agt att tta gta gaa aaa tgg atg gat act tta aaa gat aat cct tat   768
Ser Ile Leu Val Glu Lys Trp Met Asp Thr Leu Lys Asp Asn Pro Tyr
                245                 250                 255
ttt gtt aat gct ttc tct att aat att aaa act aat tta gct aaa gaa   816
Phe Val Asn Ala Phe Ser Ile Asn Ile Lys Thr Asn Leu Ala Lys Glu
            260                 265                 270
```

aag att ttt act aat att tat aat aaa tta tat agt gat tat aaa att 864
Lys Ile Phe Thr Asn Ile Tyr Asn Lys Leu Tyr Ser Asp Tyr Lys Ile
275 280 285

aat caa att aat aat cat atc cct tat tat aat tat tta aaa att aat 912
Asn Gln Ile Asn Asn His Ile Pro Tyr Tyr Asn Tyr Leu Lys Ile Asn
290 295 300

aat aaa tta cct att aaa aat att atg gat att aaa aat aat tat tgg 960
Asn Lys Leu Pro Ile Lys Asn Ile Met Asp Ile Lys Asn Asn Tyr Trp
305 310 315 320

tta gct ggt ttt aca gct gca gat ggt tct ttt tta tca tct atg tat 1008
Leu Ala Gly Phe Thr Ala Ala Asp Gly Ser Phe Leu Ser Ser Met Tyr
325 330 335

aat cct aaa gat aca tta tta ttt aaa aat atg aga cct agt tat gtt 1056
Asn Pro Lys Asp Thr Leu Leu Phe Lys Asn Met Arg Pro Ser Tyr Val
340 345 350

att tca caa gtt gaa aca cgt aaa gaa tta att tat tta att caa gaa 1104
Ile Ser Gln Val Glu Thr Arg Lys Glu Leu Ile Tyr Leu Ile Gln Glu
355 360 365

tct ttt gat tta tct att tct aat gtt aaa aaa gtt ggt aat aga aaa 1152
Ser Phe Asp Leu Ser Ile Ser Asn Val Lys Lys Val Gly Asn Arg Lys
370 375 380

tta aaa gat ttt aaa tta ttt acc aga act act gat gaa tta atg aaa 1200
Leu Lys Asp Phe Lys Leu Phe Thr Arg Thr Thr Asp Glu Leu Met Lys
385 390 395 400

ttt att tat tat ttt gat aaa ttt tta cct tta cat gat aat aaa caa 1248
Phe Ile Tyr Tyr Phe Asp Lys Phe Leu Pro Leu His Asp Asn Lys Gln

405 410 415

ttt aat tat att aaa ttt aga ttt aat act ttt att aaa tca tat aat 1296
Phe Asn Tyr Ile Lys Phe Arg Phe Asn Thr Phe Ile Lys Ser Tyr Asn
420 425 430

tgg aat aat aga gta ttt ggt tta gta tta tct gaa tat atc aat aat 1344
Trp Asn Asn Arg Val Phe Gly Leu Val Leu Ser Glu Tyr Ile Asn Asn
435 440 445

att aaa att gat aat tat gat tat tat tat tat aat aaa tat att aat 1392
Ile Lys Ile Asp Asn Tyr Asp Tyr Tyr Tyr Tyr Asn Lys Tyr Ile Asn
450 455 460

atg cat aat gca cgt aaa cct aaa gga tac att aaa taa 1431
Met His Asn Ala Arg Lys Pro Lys Gly Tyr Ile Lys
465 470 475

<210> 3
<211> 476
<212> PRT
<213> Saccharomyces cerevisiae

<400> 3
Met Lys Lys Gln Asn Leu Asn Ser Ile Leu Leu Met Tyr Ile Asn Tyr
1 5 10 15
Ile Ile Asn Tyr Phe Asn Asn Ile His Lys Asn Gln Leu Lys Lys Asp
20 25 30
Trp Ile Met Glu Tyr Glu Tyr Met Tyr Lys Phe Leu Met Asn Asn Met
35 40 45

Thr Cys Phe Ile Lys Trp Asp Asn Asn Lys Ile Leu Leu Leu Leu Asp
50 55 60
Met Tyr Tyr Asn Val Leu Tyr Asn Tyr His Lys Gln Arg Thr Pro Met
65 70 75 80
Ser Asn Lys Arg Leu Met Asn Ser Lys Asn Ile Met Asp Tyr Lys Leu
85 90 95
Leu Tyr Thr Tyr Phe Tyr Ile Leu Asn Lys Met Lys Met Glu Met Asp
100 105 110
Asn Tyr Asn Asn Asn Asn Asn Asn Ile Ser Leu Lys Tyr Asn Glu Leu
115 120 125
Leu Lys Asn Ile Met Asn Asn Leu Asn Tyr Lys Thr Ser Asn Ile Glu
130 135 140
Thr Asn Leu Ser Asn Asn Phe Tyr Leu Met Asp Lys Tyr Leu Ile Asn
145 150 155 160
Lys Tyr Met Lys Tyr Leu Asp Met Leu Asn Met Ile Pro Asn Asn Tyr
165 170 175
Met Phe Asn Asn Ile Asn Tyr Lys Gly Lys Leu Asn Ile Lys Thr Val
180 185 190
Leu Asp Leu Asn Asn Asn Glu Phe Tyr Asp Tyr Leu Ser Gly Leu Ile
195 200 205
Glu Gly Asp Gly Tyr Ile Gly Pro Gly Gly Ile Thr Ile Thr Asn His
210 215 220
Ala Asn Asp Val Leu Asn Thr Ile Phe Ile Asn Lys Arg Ile Lys Asn
225 230 235 240
Ser Ile Leu Val Glu Lys Trp Met Asp Thr Leu Lys Asp Asn Pro Tyr
245 250 255

Phe Val Asn Ala Phe Ser Ile Asn Ile Lys Thr Asn Leu Ala Lys Glu
260 265 270
Lys Ile Phe Thr Asn Ile Tyr Asn Lys Leu Tyr Ser Asp Tyr Lys Ile
275 280 285
Asn Gln Ile Asn Asn His Ile Pro Tyr Tyr Asn Tyr Leu Lys Ile Asn
290 295 300
Asn Lys Leu Pro Ile Lys Asn Ile Met Asp Ile Lys Asn Asn Tyr Trp
305 310 315 320
Leu Ala Gly Phe Thr Ala Ala Asp Gly Ser Phe Leu Ser Ser Met Tyr
325 330 335
Asn Pro Lys Asp Thr Leu Leu Phe Lys Asn Met Arg Pro Ser Tyr Val
340 345 350
Ile Ser Gln Val Glu Thr Arg Lys Glu Leu Ile Tyr Leu Ile Gln Glu
355 360 365
Ser Phe Asp Leu Ser Ile Ser Asn Val Lys Lys Val Gly Asn Arg Lys
370 375 380
Leu Lys Asp Phe Lys Leu Phe Thr Arg Thr Thr Asp Glu Leu Met Lys
385 390 395 400
Phe Ile Tyr Tyr Phe Asp Lys Phe Leu Pro Leu His Asp Asn Lys Gln
405 410 415
Phe Asn Tyr Ile Lys Phe Arg Phe Asn Thr Phe Ile Lys Ser Tyr Asn
420 425 430
Trp Asn Asn Arg Val Phe Gly Leu Val Leu Ser Glu Tyr Ile Asn Asn
435 440 445
Ile Lys Ile Asp Asn Tyr Asp Tyr Tyr Tyr Tyr Asn Lys Tyr Ile Asn
450 455 460

Met His Asn Ala Arg Lys Pro Lys Gly Tyr Ile Lys
465 470 475

<210> 4

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 4

ccggatccat gaaaaaac 18

<210> 5

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 5

gggtcgactt atttaatgta tcc 23

<210> 6

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 6

aaaagactgg attatagaa 19

<210> 7

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 7

tgaatatatg tataaattt 19

<210> 8

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 8

tattaaatgg gataataat 19

<210> 9

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 9

tattagatat gtattataat g 21

<210> 10

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 10

tacacctatg tctaataaa 19

<210> 11

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 11

aaaatattat ggattataaa 20

<210> 12

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 12

ttttatattt taaataaaat gaaaatggaa atggataatt ataataataa ta 52

<210> 13

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 13

aaaatattat gaataattta a 21

<210> 14

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 14

actatctaat attgaaacta atttatctaa taattt 36

<210> 15

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 15

ttatttaatg gataaatat 19

<210> 16

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 16

ataaatatat gaaatattta g 21

<210> 17

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 17

ataattatat gtttaataat a 21

<210> 18

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 18

ggaggtatta caattactaa tcatgctaat gat 33

<210> 19

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 19

ttttagtaga aaaatggatg gatactttaa aagata 36

<210> 20

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 20

agctaaagaa aagattttta ctaatattta taataatta 39

<210> 21

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 21

aaatattatg gatattaaa 19

<210> 22

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 22

taattattgg ttatctgg 18

<210> 23

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 23

atcatctatg tataatcct 19

<210> 24

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 24

ttaaaaatat gagacctag 19

<210> 25

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 25

gatgaattaa tgaaatttat tta 23

<210> 26

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 26

attaaattta gatttaatac tttattaaa tcatataat 39

<210> 27

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 27

tataataaat atattaatat gcataatgca cgtaaacc 38

<210> 28

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 28

taaattttta ataaataata tgacttgttt tattaaatgg ga 42

<210> 29

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 29

aagattaatg aattcaaaa 19

<210> 30

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 30

gattataaat tattatatac ttatttttat attttaaat 39

<210> 31

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 31

gaataattta aattataaaa cttctaatat tgaaacta 38

<210> 32

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 32

ttctctatta atattaaaac taatttagct aaagaaa 37

<210> 33

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 33

aaattattta ccagaactac tgatgaatta atgaaatt 38

<210> 34

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 34

catataattg gaataataga 20

<210> 35

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 35

aatttttaat gaataatatg 20

<210> 36

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 36

tttagatatg ttaaatatg 19

<210> 37

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 37

atatgttaaa tatgattcct aataa 25

<210> 38

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:Synthetic DNA

<400> 38

ctggattatg gaatatgaat 20

Claims

1. An endonuclease capable of recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof.

2. A recombinant protein of either (a) or (b):

   (a) a protein comprising the amino acid sequence represented by SEQ ID NO:3; or
   (b) a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the protein comprising at least one deletion, substitution or addition of amino acid in the amino acid sequence represented by SEQ ID NO: 3.

3. A gene encoding the recombinant protein of either (a) or (b):

   (a) a protein comprising the amino acid sequence represented by SEQ ID NO:3; or
   (b) a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the protein comprising at least one deletion, substitution or addition of amino acid in the amino acid sequence represented by SEQ ID NO: 3.

4. A gene containing DNA of either (c) or (d):

   (c) DNA comprising the nucleotide sequence represented by SEQ ID NO:2; or
   (d) DNA encoding a protein having an endonuclease activity for recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof, the DNA being capable of hybridizing with DNA which comprises the nucleotide sequence represented by SEQ ID NO:2 under stringent conditions.

5. A recombinant vector comprising the gene of claim 3 or 4.

6. A transformant comprising the recombinant vector of claim 5.

7. A process for producing the endonuclease, comprising the steps of:

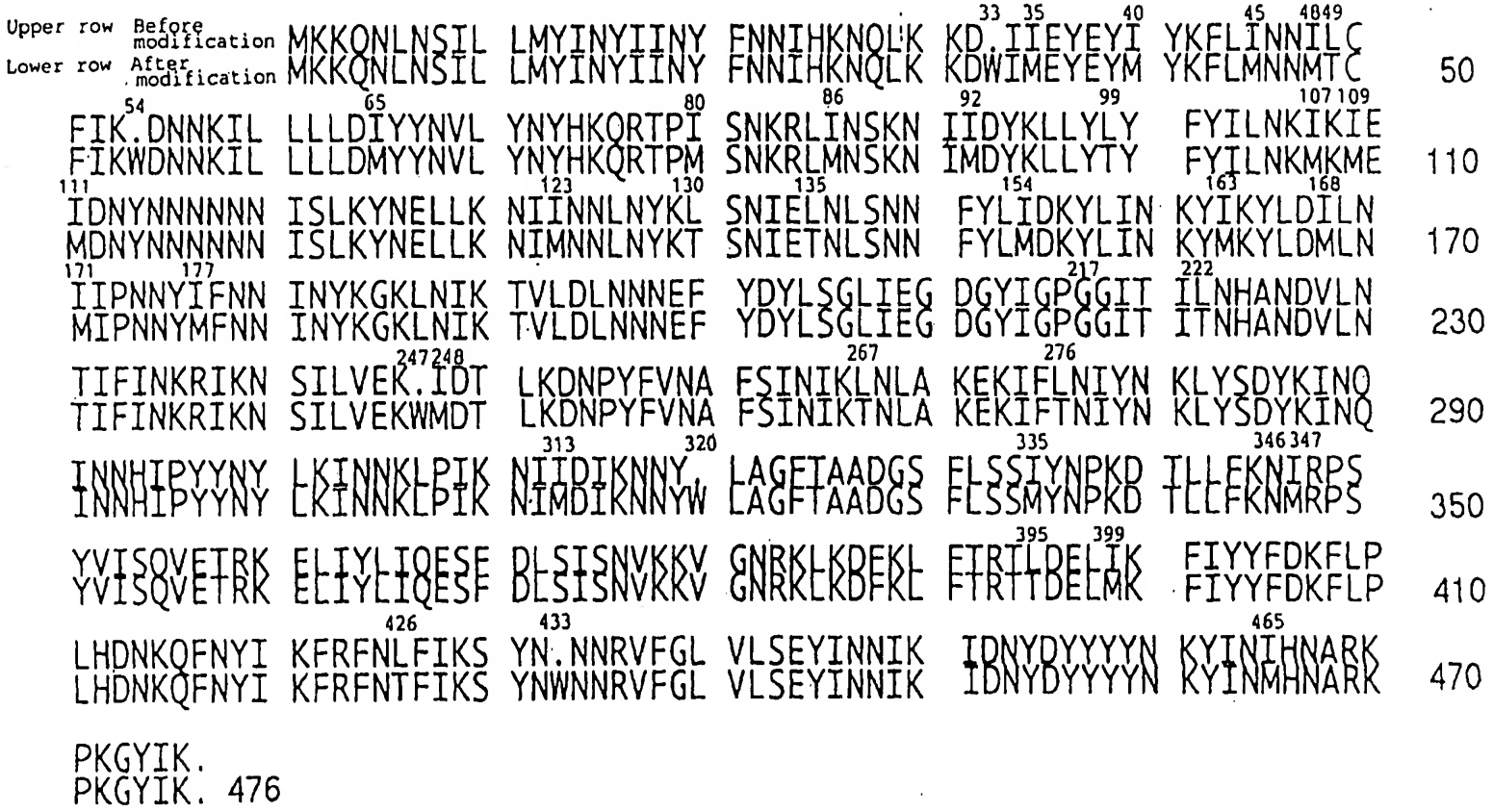
   culturing the transformant of claim 6; and
   recovering from the culture an endonuclease capable of recognizing the nucleotide sequence: GCCCAGACATATCCCTGAATGATACC or a fragment thereof.

8. An endonuclease produced by the method of claim 7.

9. Kit comprising:

   (a) the endonuclease of claim 1 or 8; and/or
   (b) the recombinant protein of claim 2, and/or
   (c) the gene of claim 3 or 4; and/or
   (d) the recombinant vector of claim 5; and/or
   (e) the transformant of claim 6.

FIG. 1



EP 0 972 836 A2

FIG. 2

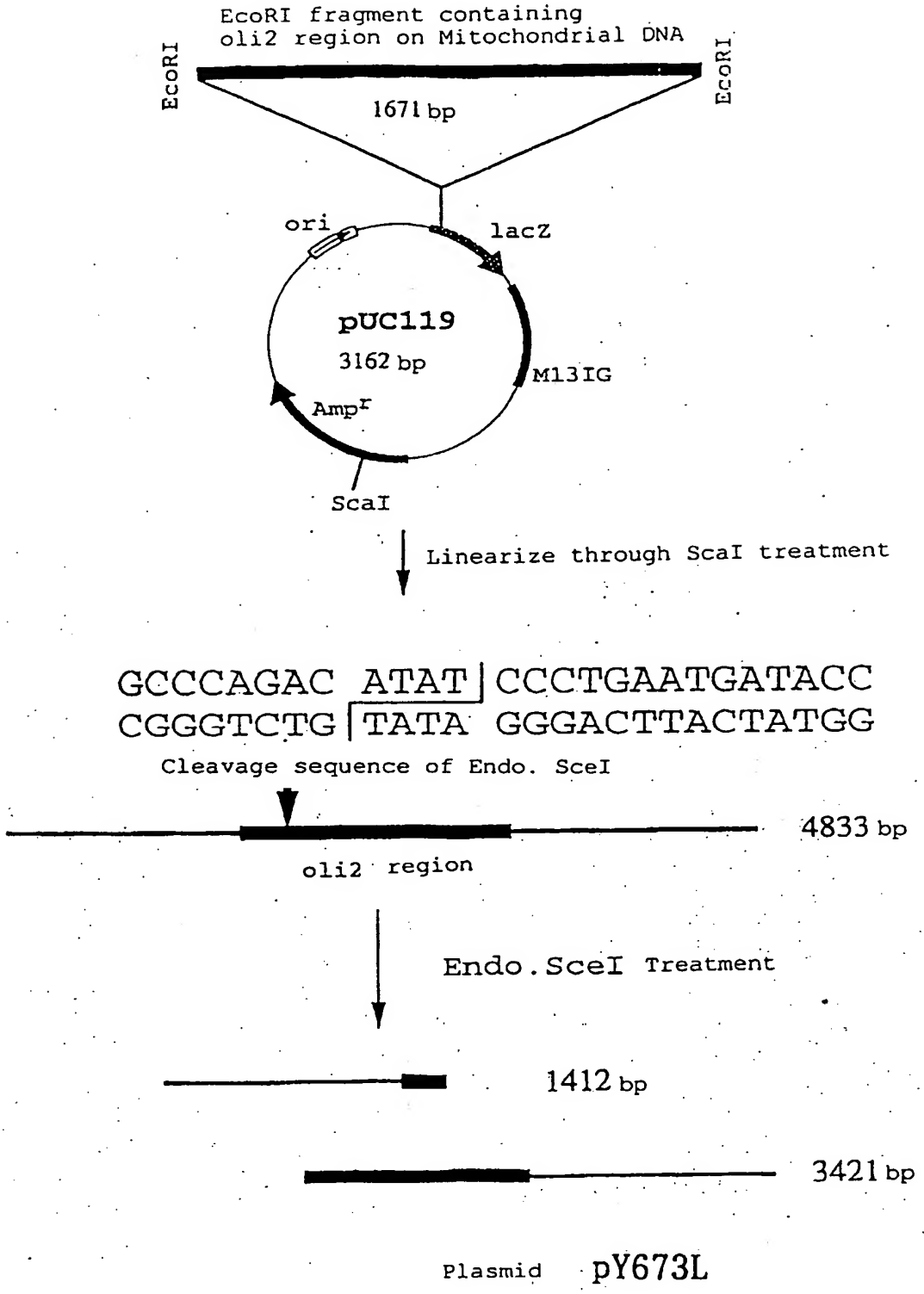
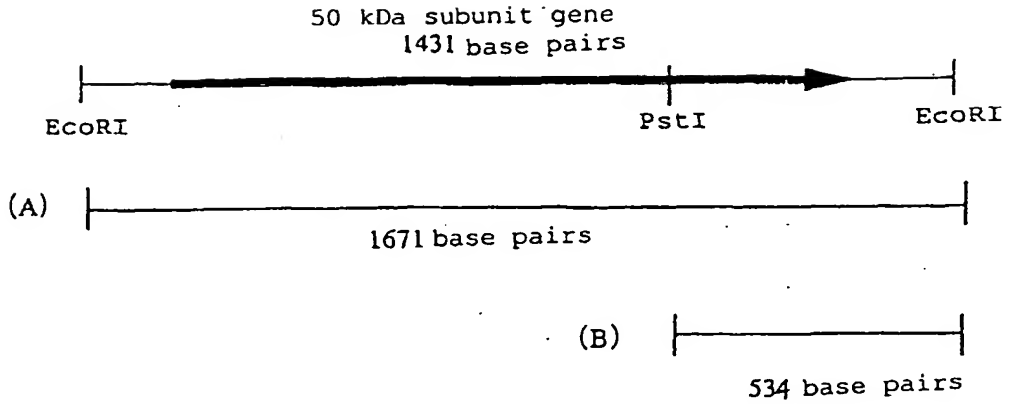
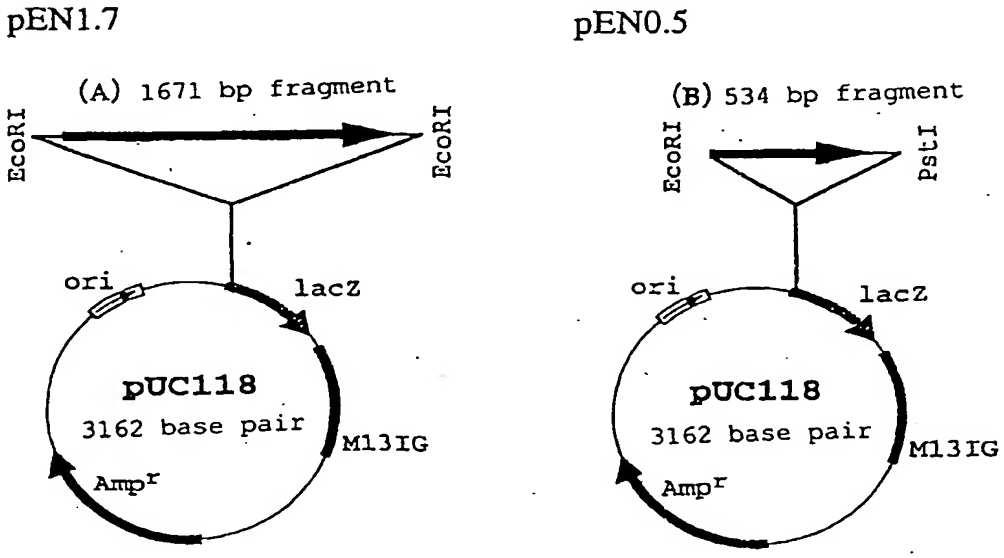

EcoRI fragment containing
oli2 region on Mitochondrial DNA
EcoRI
EcoRI
1671 bp
ori
lacZ
pUC119
3162 bp
M13IG
Amp^r
ScaI
Linearize through ScaI treatment
GCCCAGAC ATAT CCCTGAATGATACC
CGGGTCTG TATA GGGACTTACTATGG
Cleavage sequence of Endo. SceI
4833 bp
oli2 region
Endo.SceI Treatment
1412 bp
3421 bp
Plasmid pY673L

FIG. 3



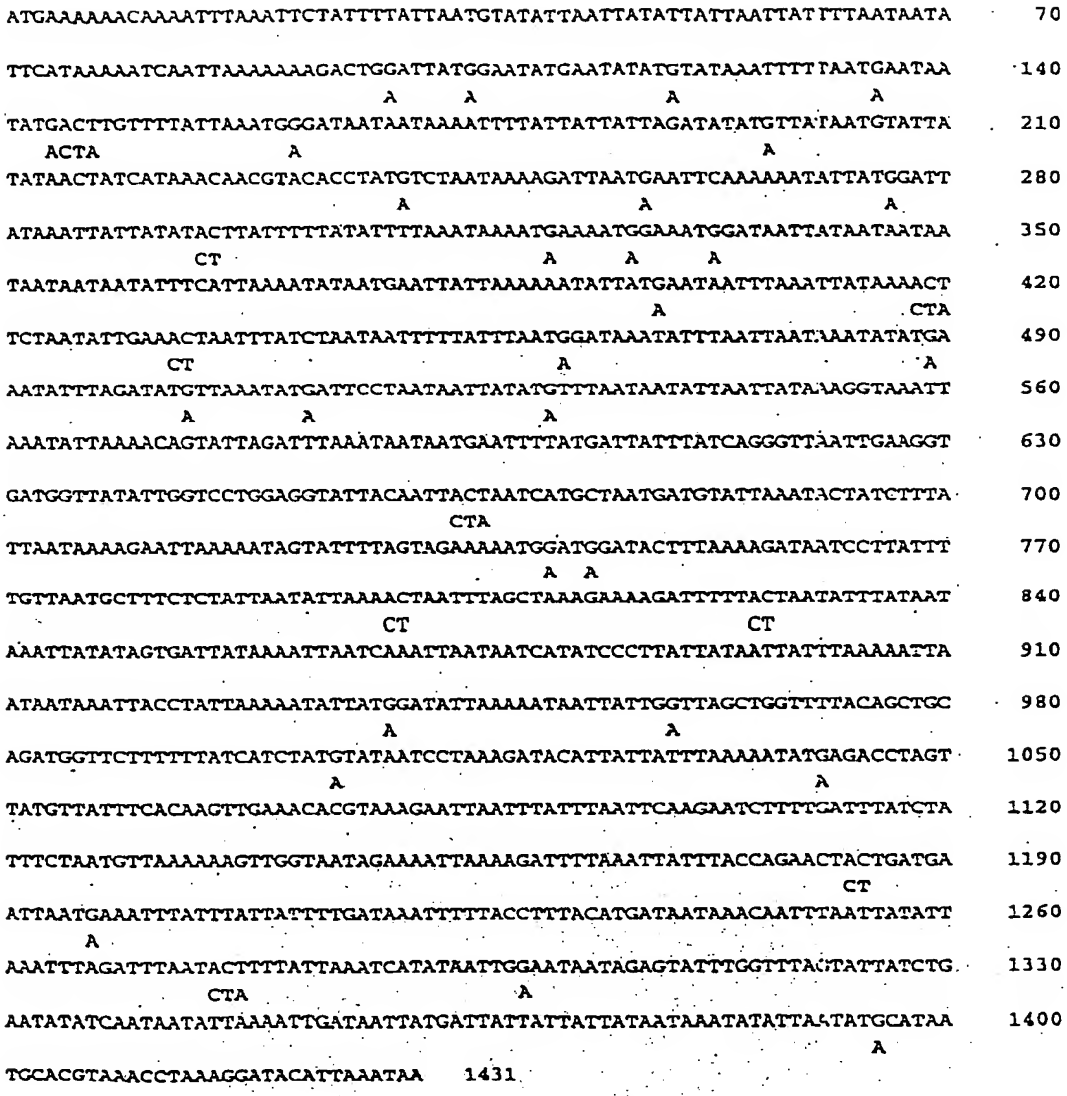


(A) Cloning the 1671 bp fragment into EcoRI site of pUC118

(B) Cloning the 534 bp fragment into EcoRI/PstI site of pUC118

Plasmids pEN1.7 and pEN0.5

FIG. 4



Nucleotide sequence of 50 kDa subunit gene which has been modified to conform to universal code (nucleotides beneath the sequence represent those before the substitution)

Sequence-specific endonuclease
activity of Endo-SceI 50 kDa subunit

FIG. 5 A Substrate pY673L

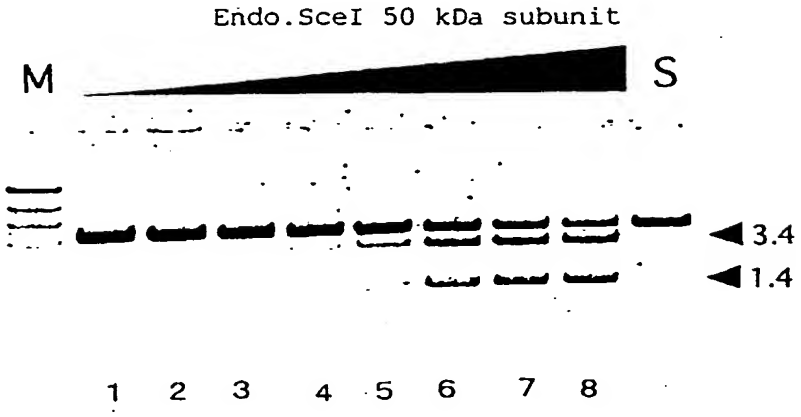
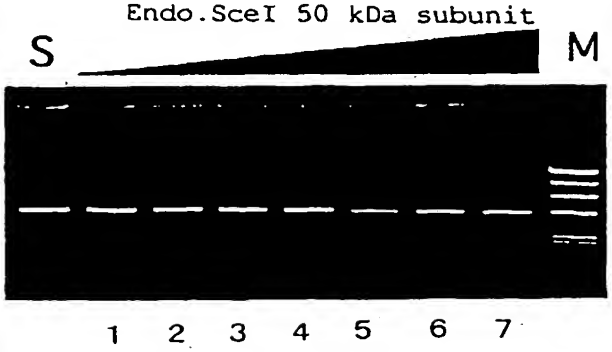


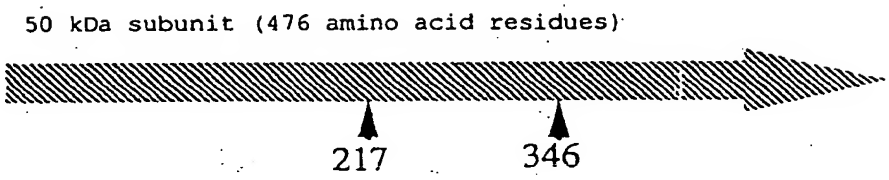
FIG. 5 B Substrate pBR322



M ; λ DNA Hind III Marker
S ; substrate DNA only

50 kDa subunit from Saccharomyces uvarum

FIG. 6A



	217	346
S.cerevisiae Endo.*Sce*I	Gly	Asn
S.uvarum Endo.*Suv*I	Lys	Asp

FIG. 6 B

35 GTTATATTGGTCCTAAAGGTATTACAATTA
(GG)

36 ATTATTTAAAGATATGAGA
(A)

Sequence-specific endonuclease
activity of Endo.SuvI 50 kDa subunit

FIG. 7A

Substrate pY673L

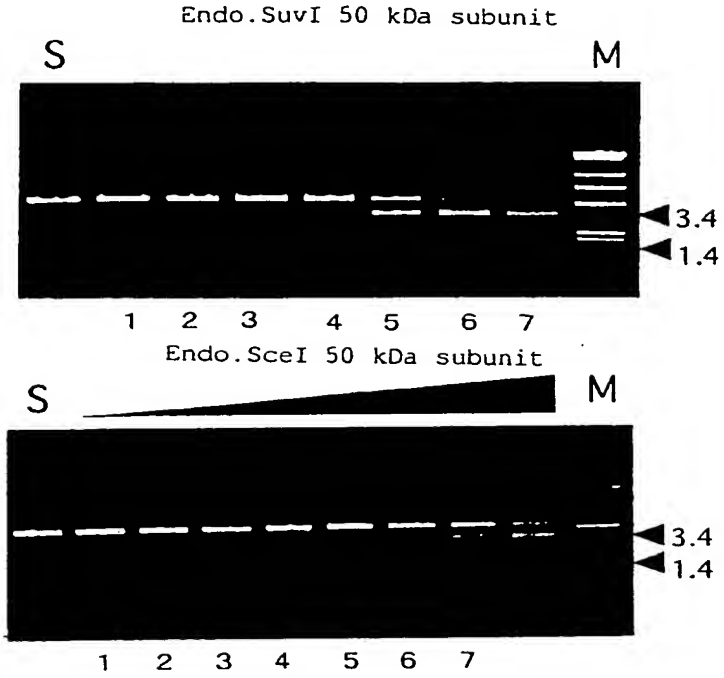
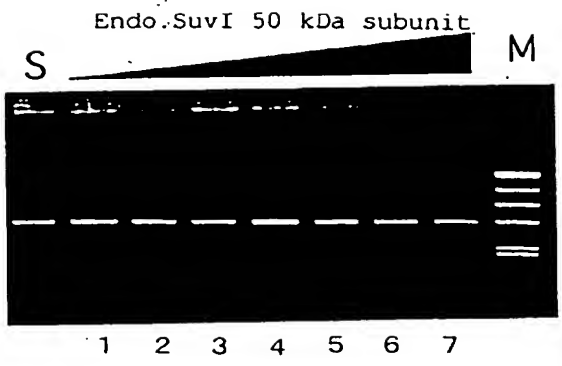


FIG. 7 B

Substrate pBR322



M ; λ DNA HindIII Marker
S ; substrate DNA only